

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/426,509ADATE: 07/10/97  
TIME: 14:19:53

INPUT SET: S18865.raw

This Raw Listing contains the General  
Information Section and up to the first 5 pages.

# 19  
7/22/97

## SEQUENCE LISTING

ENTERED

## (1) General Information

- (i) APPLICANT: Ullrich,, Axel  
Gishizsky,, Mikhail  
Sures,, Irman G.
- (ii) TITLE OF THE INVENTION: NOVEL MEGAKARYOCYTIC PROTEIN  
TYROSINE KINASES
- (iii) NUMBER OF SEQUENCES: 21
- (iv) CORRESPONDENCE ADDRESS:  
(A) ADDRESSEE: Pennie & Edmonds  
(B) STREET: 1155 Avenue of the Americas  
(C) CITY: New York,  
(D) STATE: NY  
(E) COUNTRY: USA  
(F) ZIP: 10036-2711
- (v) COMPUTER READABLE FORM:  
(A) MEDIUM TYPE: Diskette  
(B) COMPUTER: IBM Compatible  
(C) OPERATING SYSTEM: DOS  
(D) SOFTWARE: FastSEQ Version 2.0
- (vi) CURRENT APPLICATION DATA:  
(A) APPLICATION NUMBER: 08/426-509  
(B) FILING DATE: 21-APR-1995  
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:  
(A) APPLICATION NUMBER: 08/232,545  
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:  
(A) NAME: Coruzzi, Laura A  
(B) REGISTRATION NUMBER: 30,742  
(C) REFERENCE/DOCKET NUMBER: 7683-0074-999
- (ix) TELECOMMUNICATION INFORMATION:  
(A) TELEPHONE: 212-790-9090  
(B) TELEFAX: 212-869-9741

--&gt; OK

RAW SEQUENCE LISTING  
PATENT APPLICATION US/08/426,509ADATE: 07/10/97  
TIME: 14:19:55

INPUT SET: S18865.raw

47 (C) TELEX: 66141 PENNIE

48

49

50 (2) INFORMATION FOR SEQ ID NO:1:

51

52 (i) SEQUENCE CHARACTERISTICS:

53 (A) LENGTH: 2000 base pairs

54 (B) TYPE: nucleic acid

55 (C) STRANDEDNESS: unknown

56 (D) TOPOLOGY: unknown

57

58

59 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

60

61	CTCGCTCCAA	GTTGTGCAGC	CGGGACCGCC	TCGGGGTGTG	CAGCCGGCTC	GCGGAGGCC	60
62	TCCTGGGGGC	GGGCGCGGGG	CGGCTCGGGG	GCGCCCCCTG	AGCAGAAAAC	AGGAAGAACC	120
63	AGGCTCGGTC	CAGTGGCACC	CAGCTCCCTA	CCTCCTGTGC	CAGCCGCCTG	GCCTGTGGCA	180
64	GGCCATTCCC	AGCGTCCCCG	ACTGTGACCA	CTTGCTCAGT	GTGCCTCTCA	CCTGCCTCAG	240
65	TTTCCCTCTG	GGGGGCGATG	GCGGGGCGAG	GCTCTCTGGT	TTCTTGCGCG	GCATTTTCACG	300
66	GCTGTGATTG	TGCTGAGGAA	CTTCCCCGGG	TGAGCCCCCG	CTTCCTCCGA	GCCTGGCACC	360
67	CCCCTCCCGT	CTCAGCCAGG	ATGCCAACGA	GGCGCTGGGC	CCCGGGCACC	CAGTGTATCA	420
68	CCAAATGCGA	GCACACCCGC	CCCAAGCCAG	GGGAGCTGGC	CTTCCGCAAG	GGCGACGTGG	480
69	TCACCATCCT	GGAGGCCTGC	GAGAACAAGA	GCTGGTACCG	CGTCAAGCAC	CACACCAGTG	540
70	GACAGGAGGG	GCTGCTGGCA	GCTGGGGCGC	TGCGGGAGCG	GGAGGCCCTC	TCCGCAGACC	600
71	CCAAGCTCAG	CCTCATGCCG	TGGTTCCACG	GGAAGATCTC	GGGCCAGGAG	GCTGTCCAGC	660
72	AGCTGCAGCC	TCCCCAGGAT	GGGCTGTTCC	TGGTGCGGGA	GTCCGCGCGC	CACCCCGGCG	720
73	ACTACGTCTT	GTGCGTGAGC	TTTGGCCGCG	ACGTCAATCA	CTACCGCGTG	CTGCACCGCG	780
74	ACGGCCACCT	CACAATCGAT	GAGGCCGTGT	TCTTCTGCAA	CCTCATGGAC	ATGGTGGAGC	840
75	ATTACAGCAA	GGACAAGGGC	GCTATCTGCA	CCAAGCTGGT	GAGACCAAAG	CGGAAACACG	900
76	GGACCAAGTC	GGCCGAGGAG	GAGCTGGCCA	GGGCGGGCTG	GTTACTGAAC	CTGCAGCATT	960
77	TGACATTGGG	AGCACAGATC	GGAGAGGGAG	AGTTTGGAGC	TGTCTGTCAG	GGTGAGTACC	1020
78	TGGGGCAAAA	GGTGGCCGTG	AAGAATATCA	AGTGTGATGT	GACAGCCCAG	GCCTTCCTGG	1080
79	ACGAGACGGC	CGTCATGACG	AAGATGCAAC	ACGAGAACCT	GGTGCGTCTC	CTGGGCGTGA	1140
80	TCCTGCACCA	GGGGCTGTAC	ATTGTCTATG	AGCACGTGAG	CAAGGGCAAC	CTGGTGAAC	1200
81	TTCTGCGGAC	CCGGGGTCGA	GCCCTCGTGA	ACACCGCTCA	GCTCCTGCAG	TTTTCTCTGC	1260
82	ACGTGGCCGA	GGGCATGGAG	TACCTGGAGA	GCAAGAAGCT	TGTGCACCGC	GACCTGGCCG	1320
83	CCCGCAACAT	CCTGGTCTCA	GAGGACCTGG	TGGCCAAGGT	CAGCGACTTT	GGCCTGGCCA	1380
84	AAGCCGAGCG	GAAGGGGCTA	GAATCAAGCC	GGCTGCCCCG	CAAGTGGACG	GCGCCCGAGG	1440
85	CTCTCAAACA	CGGGAAGTTC	ACCAGCAAGT	CGGATGTCTG	GAGTTTGGGG	GTGCTGCTCT	1500
86	GGGAGGTCTT	CTCATATGGA	CGGGCTCCGT	ACCCTAAAAT	GTCCTGAAA	GAGGTGTCGG	1560
87	AGGCCGTGGA	GAAGGGGTAC	CGCATGGAAC	CCCCCGAGGG	CTGTCCAGGC	CCCGTGCACG	1620
88	TCCTCATGAG	CAGCTGCTGG	GAGGCAGAGC	CCGCCCCCGG	GCCACCCCTT	CGCAAACCTG	1680
89	CCGAGAAGCT	GGCCCGGGAG	CTACGCAAGT	CAGGTGCCCC	AGCCTCCGTC	TCAGGGCAGG	1740
90	ACGCCGACGG	CTCCACCTCG	CCCCGAAGCC	AGGAGCCCTG	ACCCACCCCG	GTGGGGCCCT	1800
91	TGGCCCCAGA	GGACCGAGAG	AGTGGAGAGT	GCGGCGTGGG	GGCACTGACC	AGGCCCAAGG	1860
92	AGGGTCCAGG	CGGGCAAGTC	ATCCTCCTGG	TGCCCCACAGC	AGGGGCTGGC	CCACGTAGGG	1920
93	GGCTCTGGGC	GGCCCGTGGG	CACCCAGAC	CTGCGAAGGA	TGATCGCCCC	ATAAAGACGG	1980
94	ATTCTAAGGA	CTCTAAAAAA					2000

95

96 (2) INFORMATION FOR SEQ ID NO:2:

97

98 (i) SEQUENCE CHARACTERISTICS:

99 (A) LENGTH: 507 amino acids

# RAW SEQUENCE LISTING

## PATENT APPLICATION US/08/426,509A

DATE: 07/10/97

TIME: 14:19:57

INPUT SET: S18865.raw

```

100      (B) TYPE: amino acid
101      (C) STRANDEDNESS: unknown
102      (D) TOPOLOGY: unknown
103
104      (ii) MOLECULE TYPE: None
105
106      (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:
107
108      Met Ala Gly Arg Gly Ser Leu Val Ser Trp Arg Ala Phe His Gly Cys
109      1      5      10      15
110      Asp Ser Ala Glu Glu Leu Pro Arg Val Ser Pro Arg Phe Leu Arg Ala
111      20      25      30
112      Trp His Pro Pro Val Ser Ala Arg Met Pro Thr Arg Arg Trp Ala
113      35      40      45
114      Pro Gly Thr Gln Cys Ile Thr Lys Cys Glu His Thr Arg Pro Lys Pro
115      50      55      60
116      Gly Glu Leu Ala Phe Arg Lys Gly Asp Val Val Thr Ile Leu Glu Ala
117      65      70      75      80
118      Cys Glu Asn Lys Ser Trp Tyr Arg Val Lys His His Thr Ser Gly Gln
119      85      90      95
120      Glu Gly Leu Leu Ala Ala Gly Ala Leu Arg Glu Arg Glu Ala Leu Ser
121      100      105      110
122      Ala Asp Pro Lys Leu Ser Leu Met Pro Trp Phe His Gly Lys Ile Ser
123      115      120      125
124      Gly Gln Glu Ala Val Gln Gln Leu Gln Pro Pro Glu Asp Gly Leu Phe
125      130      135      140
126      Leu Val Arg Glu Ser Ala Arg His Pro Gly Asp Tyr Val Leu Cys Val
127      145      150      155      160
128      Ser Phe Gly Arg Asp Val Ile His Tyr Arg Val Leu His Arg Asp Gly
129      165      170      175
130      His Leu Thr Ile Asp Glu Ala Val Phe Phe Cys Asn Leu Met Asp Met
131      180      185      190
132      Val Glu His Tyr Ser Lys Asp Lys Gly Ala Ile Cys Thr Lys Leu Val
133      195      200      205
134      Arg Pro Lys Arg Lys His Gly Thr Lys Ser Ala Glu Glu Glu Leu Ala
135      210      215      220
136      Arg Ala Gly Trp Leu Leu Asn Leu Gln His Leu Thr Leu Gly Ala Gln
137      225      230      235      240
138      Ile Gly Glu Gly Glu Phe Gly Ala Val Leu Gln Gly Glu Tyr Leu Gly
139      245      250      255
140      Gln Lys Val Ala Val Lys Asn Ile Lys Cys Asp Val Thr Ala Gln Ala
141      260      265      270
142      Phe Leu Asp Glu Thr Ala Val Met Thr Lys Met Gln His Glu Asn Leu
143      275      280      285
144      Val Arg Leu Leu Gly Val Ile Leu His Gln Gly Leu Tyr Ile Val Met
145      290      295      300
146      Glu His Val Ser Lys Gly Asn Leu Val Asn Phe Leu Arg Thr Arg Gly
147      305      310      315      320
148      Arg Ala Leu Val Asn Thr Ala Gln Leu Leu Gln Phe Ser Leu His Val
149      325      330      335
150      Ala Glu Gly Met Glu Tyr Leu Glu Ser Lys Lys Leu Val His Arg Asp
151      340      345      350
152      Leu Ala Ala Arg Asn Ile Leu Val Ser Glu Asp Leu Val Ala Lys Val

```

# RAW SEQUENCE LISTING PATENT APPLICATION US/08/426,509A

DATE: 07/10/97  
TIME: 14:19:59

INPUT SET: S18865.raw

```

153          355          360          365
154 Ser Asp Phe Gly Leu Ala Lys Ala Glu Arg Lys Gly Leu Asp Ser Ser
155          370          375          380
156 Arg Leu Pro Val Lys Trp Thr Ala Pro Glu Ala Leu Lys His Gly Lys
157          385          390          395          400
158 Phe Thr Ser Lys Ser Asp Val Trp Ser Phe Gly Val Leu Leu Trp Glu
159          405          410          415
160 Val Phe Ser Tyr Gly Arg Ala Pro Tyr Pro Lys Met Ser Leu Lys Glu
161          420          425          430
162 Val Ser Glu Ala Val Glu Lys Gly Tyr Arg Met Glu Pro Pro Glu Gly
163          435          440          445
164 Cys Pro Gly Pro Val His Val Leu Met Ser Ser Cys Trp Glu Ala Glu
165          450          455          460
166 Pro Ala Arg Arg Pro Pro Phe Arg Lys Leu Ala Glu Lys Leu Ala Arg
167          465          470          475          480
168 Glu Leu Arg Ser Ala Gly Ala Pro Ala Ser Val Ser Gly Gln Asp Ala
169          485          490          495
170 Asp Gly Ser Thr Ser Pro Arg Ser Gln Glu Pro
171          500          505

```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2500 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```

184 CCGCTTTTTG CTTAGAGCTT GAGAGTCAAA GTTAAGGACC CACATGTATA CTTCGGCTCT 60
185 AGCGAGTCTA AGGATGATAA TATGGATACA AAATCTATTC TAGAAGAACT TCTTCTCAAA 120
186 AGATCACAGC AAAAGAAGAA AATGTCACCA AATAATTACA AAGAACGGCT TTTTGTTTTG 180
187 ACCAAAACAA ACCTTTCCTA CTATGAATAT GACAAAATGA AAAGGGGCAG CAGAAAAGGA 240
188 TCCATTGAAA TTAAGAAAAAT CAGATGTGTG GAGAAAAGTAA ATCTCGAGGA GCAGACGCCT 300
189 GTAGAGAGAC AGTACCCATT TCAGATTGTC TATAAAGATG GGCTTCTCTA TGTCTATGCA 360
190 TCAAATGAAG AGAGCCGAAG TCAGTGTTTG AAAGCATTAC AAAAAGAGAT AAGGGGTAAC 420
191 CCCCACCTGC TGGTCAAGTA CCATAGTGGG TTCTTCGTGG ACGGGAAGTT CCTGTGTTGC 480
192 CAGCAGAGCT GTAAAGCAGC CCCAGGATGT ACCCTCTGGG AAGCATATGC TAATCTGCAT 540
193 ACTGCAGTCA ATGAAGAGAA ACACAGAGTT CCCACCTTCC CAGACAGAGT GCTGAAGATA 600
194 CCTCGGGCAG TTCCTGTTCT CAAAATGGAT GCACCATCTT CAAGTACCAC TCTAGCCCAA 660
195 TATGACAACG AATCAAAGAA AAACATATGGC TCCCAGCCAC CATCTTCAAG TACCAGTCTA 720
196 GCGCAATATG ACAGCAACTC AAAGAAAATC TATGGCTCCC AGCCAAACTT CAACATGCAG 780
197 TATATTCCAA GGGAAGACTT CCCTGACTGG TGGCAAGTAA GAAAACTGAA AAGTAGCAGC 840
198 AGCAGTGAAG ATGTTGCAAG CCGTAACCAA AAAGAAAGAA ATGTGAATCA CACCACCTCA 900
199 AAGATTTTCAT GGGAAATCCC TGAGTCAAGT TCATCTGAAG AAGAGGAAAA CCTGGATCAT 960
200 TATGACTGGT TTGCTGGTAA CATCTCCAGA TCACAATCTG AACAGTTACT CAGACAAAAG 1020
201 GGAAAAGAAG GAGCATTTAT GGTTAGAAAT TCGAGCCAAG TGGGAATGTA CAGAGTGTCC 1080
202 TTATTTAGTA AGGCTGTGAA TGATAAAAAA GGAAGTGTCA AACATTACCA CGTGCATACA 1140
203 AATGCTGAGA ACAAATTATA CCTGGCAGAA AACTACTGTT TTGATTCCAT TCCAAAGCTT 1200
204 ATTCATTATC ATCAACACAA TTCAGCAGGC ATGATCACAC GGCTCCGCCA CCCTGTGTCA 1260
205 ACAAAGGCCA ACAAGGTCCC CGACTCTGTG TCCCTGGGAA ATGGAATCTG GGAAGTGAAG 1320

```

# RAW SEQUENCE LISTING PATENT APPLICATION US/08/426,509A

DATE: 07/10/97  
TIME: 14:20:02

INPUT SET: S18865.raw

206	AGAGAAGAGA	TTACCTTGTT	GAAGGAGCTG	GGAAGTGGCC	AGTTTGGAGT	GGTCCAGCTG	1380
207	GGCAAGTGGA	AGGGGCAGTA	TGATGTTGCT	GTTAAGATGA	TCAAGGAGGG	CTCCATGTCA	1440
208	GAAGATGAAT	TCCTTCAGGA	GGCCCAGACT	ATGATGAAAC	TCAGCCATCC	CAAGCTGGTT	1500
209	AAATTCTATG	GAGTGTGTTT	AAAGGAATAC	CCCATATACA	TAGTGACTGA	ATATATAAGC	1560
210	AATGGCTGCT	TGCTGAATTA	CCTGAGGAGT	CACGGAAAAG	GACTTGAACC	TTCCCAGCTC	1620
211	TTAGAAATGT	GCTACGATGT	CTGTGAAGGC	ATGGCCTTCT	TGGAGAGTCA	CCAATTCATA	1680
212	CACCGGGACT	TGGCTGCTCG	TAAGTCTTGG	GTGGACAGAG	ATCTCTGTGT	GAAAGTATCT	1740
213	GACTTTGGAA	TGACAAGGTA	TGTTCTTGAT	GACCAAGTATG	TCAGTTCAGT	CGGAACAAAG	1800
214	TTTCCAGTCA	AGTGGTCAGC	TCCAGAGGTG	TTTCATTACT	TCAAATACAG	CAGCAAGTCA	1860
215	GACGTATGGG	CATTTGGGAT	CCTGATGTGG	GAGGTGTTCA	GCCTGGGGAA	GCAGCCCTAT	1920
216	GACTTGATATG	ACAAC TCCCA	GGTGGTTCTG	AAGGTCTCCC	AGGGCCACAG	GCTTTACCGG	1980
217	CCCCACCTGG	CATCGGACAC	CATCTACCAG	ATCATGTACA	GCTGCTGGCA	CGAGCTTCCA	2040
218	GAAAAGCGTC	CCACATTTCA	GCAACTCTCT	TCTTCCATTG	AACCACTTCG	GGAAAAAGAC	2100
219	AAGCATTGAA	GAAGAAATTA	GGAGTGCTGA	TAAGAATGAA	TATAGATGCT	GGCCAGCATT	2160
220	TTCATTTCATT	TTAAGGAAAG	TAGCAAGGCA	TAATGTAATT	TAGCTAGTTT	TTAATAGTGT	2220
221	TCTCTGTATT	GTCTATTATT	TAGAAATGAA	CAAGGCAGGA	AACAAAAGAT	TCCCTTGAAA	2280
222	TTTAGGTCAA	ATTAGTAATT	TTGTTTATGC	TGCCCTTGAT	ATAACACTTT	CCAGCCTATA	2340
223	GCAGAAGCAC	ATTTTCAGAC	TGCAATATAG	AGACTGTGTT	CATGTGTAAA	GACTGAGCAG	2400
224	AACTGAAAAA	TTACTTATTG	GATATTCATT	CTTTTCTTTA	TATTGTCATT	GTCACAACAA	2460
225	TTAAATATAC	TACCAAGTAC	AAAAAAAAAA	AAAAAAAAAA			2500

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 675 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: unknown

(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: None

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

239	Met	Asp	Thr	Lys	Ser	Ile	Leu	Glu	Glu	Leu	Leu	Lys	Arg	Ser	Gln
240	1				5					10				15	
241	Gln	Lys	Lys	Lys	Met	Ser	Pro	Asn	Asn	Tyr	Lys	Glu	Arg	Leu	Phe
242				20					25				30		Val
243	Leu	Thr	Lys	Thr	Asn	Leu	Ser	Tyr	Tyr	Glu	Tyr	Asp	Lys	Met	Lys
244			35					40				45			Arg
245	Gly	Ser	Arg	Lys	Gly	Ser	Ile	Glu	Ile	Lys	Lys	Ile	Arg	Cys	Val
246		50				55						60			Glu
247	Lys	Val	Asn	Leu	Glu	Glu	Gln	Thr	Pro	Val	Glu	Arg	Gln	Tyr	Pro
248		65				70					75				80
249	Gln	Ile	Val	Tyr	Lys	Asp	Gly	Leu	Leu	Tyr	Val	Tyr	Ala	Ser	Asn
250				85					90					95	Glu
251	Glu	Ser	Arg	Ser	Gln	Trp	Leu	Lys	Ala	Leu	Gln	Lys	Glu	Ile	Arg
252				100					105					110	Gly
253	Asn	Pro	His	Leu	Leu	Val	Lys	Tyr	His	Ser	Gly	Phe	Phe	Val	Asp
254			115					120				125			Gly
255	Lys	Phe	Leu	Cys	Cys	Gln	Gln	Ser	Cys	Lys	Ala	Ala	Pro	Gly	Cys
256		130					135					140			Thr
257	Leu	Trp	Glu	Ala	Tyr	Ala	Asn	Leu	His	Thr	Ala	Val	Asn	Glu	Glu
258		145				150					155				Lys

PAGE: 1

**SEQUENCE VERIFICATION REPORT**  
**PATENT APPLICATION US/08/426,509A**

DATE: 07/10/97  
TIME: 14:20:04

*INPUT SET: S18865.raw*

Line	Error	Original Text
29	Wrong application Serial Number	(A) APPLICATION NUMBER: 08/426-509